Twentieth-Century Music 19/3, 427–443 © The Author(s), 2022. Published by Cambridge University Press
doi: 10.1017/S1478572222000287

Autoplaying, Unmuting, Attending: (Re)formatting the Twenty-First-Century Digital Sensorium

PAULA HARPER

Abstract

In this article, I analyse the implications of autoplaying video as a driver of 'audile techniques' in the 2010s digital ecosystem – in particular, techniques that respond to the realities of the separability of image and sound, even in media that contain both elements. I then examine a number of strategies through which this audio/visual split has been negotiated, monetized, and creatively bridged by consumers, creators, and corporate personnel – from the creation of new audiovisual genres and aesthetics, to the rise of particular platform pricing models, to the adoption (and, potentially, exploitation) of accessibility features. Ultimately, I seek to show how negotiations of sound and listening factor deeply into contemporary attempts to harness and monetize 'attention' as a commodity in a digital economy of platforms, advertisements, and data.

One afternoon prior to the global COVID-19 pandemic, while rushing through midtown Manhattan, I passed a billboard ad for online and digital streaming radio service Pandora. 'Running late? Keep calm and Sound On', the advertisement says. The 'Keep Calm' portion is neatly formatted in reference to the iconic British wartime propaganda posters, but the 'SOUND ON' portion leaps out in a bright colour, huge font, and a typeface designed to evoke scrawled handwriting (Figure 1).

I snapped a picture as I bustled past.

I took the picture because the tagline struck me as familiar – not from Pandora or any other corporate context, but from everyday users' internet captions, which had caught my eye a few years prior across social media platforms such as Tumblr and Twitter. These more familiar vernacular captions were ones that users tended to make on online video posts – what I have elsewhere called the 'unmute this' video microgenre – urging their fellow feed-scrollers to turn the sound on for a particular bit of video content.[1] For me, these two examples – the Pandora 'Sound On' campaign and earlier 'unmute this' captions – stand as two instances in a broader network of negotiations of a curious sensory flip, a rift in the contemporary digital sensorium, in which sound and image are fundamentally separate (or separable), and in which it is listening, rather than viewing, that is construed by designers, advertisers, content creators, and everyday users as an elective, fully attentive act.

Over the course of the 2000s and early 2010s, producers and consumers of digital audiovisual media have both effected and navigated the bifurcation of viewing and listening in a

Email: pcharper@uchicago.edu

1 Paula Harper, '"Unmute This": Captioning an (Audio)visual Microgenre', *The Soundtrack* 9/1–2 (2016).

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press



Figure 1 (Colour online) Pandora 'Keep Calm and Sound On' billboard advertisement above a New York City Subway entrance (27 December 2018).

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

milieu of ubiquitous digital devices. Responses to this split – a partitioning off of image and sound, through an ecosystem of computers and smartphones, earbuds, and digital platforms – have evolved dialectically, as advertisers and corporations react to user behaviour, and as users respond to new corporate standards and impositions.

In this article, I situate autoplaying video as a central problematic format in this assemblage. I then examine a number of strategies through which the audio/visual split of the 2010s digital ecosystem has been negotiated, monetized, and creatively bridged by consumers, creators, and corporate personnel – from the creation of new audiovisual genres and aesthetics, to the rise of particular platform pricing models, to the adoption (and, potentially, exploitation) of accessibility features. Ultimately, I seek to show how negotiations of sound and listening factor deeply into contemporary attempts to harness and monetize 'attention' as a commodity in a digital economy of platforms, advertisements, and data.

Autoplaying video's Facebook revival

A Facebook press release issued on 12 September 2013 announced initial tests of 'an easier way to watch videos on Facebook'. 'Now', the release apprised, 'when you see a video in News Feed, it comes to life and starts playing.'[2] The Facebook release also stressed the limited scope of intended participants in the sharing of such videos: performers, musicians, and everyday users would be creating and posting videos, but advertisers would *not* be. The Facebook press release touted this limitation as a means of fostering 'the best possible experience' for its (non-corporate) users. However, it acknowledged that 'over time, we'll continue to explore how to bring this to marketers in the future'.

Of course, that 'exploration' soon came to fruition in the originally intended product, and autoplaying video advertisements proliferated among other user-created and shared videos. The introduction of autoplaying video through its palatable, even appealing use by popular performers and members of Facebook's users' social circles had functioned as habituation for an ideal advertising format. By 2015, Facebook's rebranding of autoplaying video was a success. A piece from media trade blog *Digiday* declared 'Like it or not: autoplay video won', citing Facebook's implementation as the vanguard for the adoption of various video autoplay functionalities across a number of other platforms, including Twitter, Instagram, and YouTube – not to mention launching the now somewhat-infamous 'pivot to video' across the media landscape more broadly.[3] The spread of the autoplay functionality provoked novel

2 Kelly Mayes, 'An Easier Way to Watch Video', *Facebook Newsroom*, 12 September 2013, https://about.fb.com/news/2013/09/an-easier-way-to-watch-video/.

3 Eric Blattberg, 'Like It or Not, Autoplay Video Won', *Digiday*, 21 April 2015, http://digiday.com/publishers/autoplay-video-beat-regular-video-sorry-guys/. It is beyond the scope of this article, but nonetheless worth remarking on, that Facebook circulated and promoted data hugely overstating the impact and viewership of video on its platform, based on misleading metrics for what 'counted' as video views. Facebook disclosed this in 2016, but the admission by that point could not undo changes already undertaken broadly in journalism, media, advertising, and other industries, which in response to Facebook's misleading data (and overwhelming platform clout) had significantly restructured and reoriented their production efforts – oftentimes manifesting in major layoffs and other infrastructural upheaval. See Suzanne Vranica and Jack Marshall, 'Facebook Overestimated Key Video Metric for Two Years', *Wall Street*

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

kinds of user behaviour in response, and ultimately engendered the formation of new audio-visual artefacts, genres, and encounters.

Yet the success of Facebook's autoplay story happened against a backdrop of considerable hostility. Indeed, in 2013 – the year that Facebook launched its autoplay video rollout – a different *Digiday* piece had referred to autoplay as 'the most hated digital ad tactic', citing widespread industry and consumer resistance to it.[4] In the late 2000s, alongside increasingly ubiquitous mobile phone use and large-scale adoption of broadband internet access (with accompanying dramatic and ongoing increases in bandwidth), digital marketing platforms had begun offering video advertisements among their array of online advertising products. Many of these videos were highly visually intrusive, designed to block much of a website's onscreen real estate – and designed to begin to play automatically, without warning to the viewer and without providing any means for the viewer to stop or remove the ad. User response to such ad tactics was almost uniformly negative, and user-generated 'how-to' posts proliferated across internet forums and tech blogs, offering step-by-step instructions for disrupting or disabling such advertisements on specific websites or entire browsers. Implementing autoplaying video was construed not just as an annoyance, but as objectively bad internet practice – the 2008 Web Content Accessibility Guidelines (WCAG) assembled by the World Wide Web Consortium (W3C) recommended against the use of autoplay as an *in*accessible tactic, causing negative effects for a variety of users with cognitive disabilities, seizure propensities, or those using screen-readers or other additive accessibility tools.

As a simplification of this broad web of concerns, critiques tended to coalesce around a key driver of hatred for autoplaying ads – their autoplaying sound. In the *Digiday* piece, the author sets the following scene:

> You're sitting in front of your computer at work, and you're surfing Twitter. A friend shares a link which you find interesting. You click and as soon as the page opens up, your face gets embarrassingly red as everyone in your silent office looks at you. You were just served an autoplay video.[5]

This scenario demonstrates an assumed environment for *Digiday*'s demographic of web users: an open office, a bullpen or cubicles, with desktop computers' built-in speakers projecting surprising sound into the otherwise-quiet professional space. While screens might be directional, sound in this scenario has the troubling capacity to traverse boundaries such as cubicle walls – and to upset a particular normative sensory order in a white-collar corporate space.

Facebook's 2013 implementation of autoplaying video occurred, then, against a backdrop of powerful aversion to the technique. The wording of Facebook's initial release, the assurances

Journal, 22 September 2016, www.wsj.com/articles/facebook-overestimated-key-video-metric-for-two-years-1474586951; Laura Hazard Owen, 'Did Facebook's Faulty Data Push News Publishers to Make Terrible Decisions on Video?', *NiemanLab.org*, 17 October 2018, www.niemanlab.org/2018/10/did-facebooks-faulty-data-push-news-publishers-to-make-terrible-decisions-on-video/.

4 Josh Sternberg, 'The Most Hated Digital Ad Tactic', *Digiday.com*, 3 April 2013, https://digiday.com/media/the-most-hated-digital-ad-tactic/.

5 Sternberg, 'The Most Hated Digital Ad Tactic'.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

that the videos autoplaying on feeds would not be advertisements (at least, not initially) but would be music, performances, or amateur creations – these all served to push back against the extant user and tech-world feelings regarding autoplaying video. A final, crucial component to making autoplaying video a palatable standard was avoiding the embarrassing social situation depicted in the *Digiday* piece. Facebook's new autoplaying videos did not autoplay in their entirety, but separated out the parameters of image and sound – while videos' visual tracks began playing as a user scrolled past them, their audio would stay silent, until or unless the user decided to click to turn it on.

Sound on and attention in the digital sensorium

In *The Audible Past*, Jonathan Sterne theorizes 'audile techniques', or 'techniques of listening', expanding from Marcel Mauss's 'techniques of the body'.[6] For Sterne, 'audile techniques' manifest in novel situations of sensory separability, in which a variety of devices – and their application and use – make possible the isolation of sound and listening as particular practices, particular techniques. When I discuss a separability between image and sound, vision and listening online and in digital media, negotiated by corporate and vernacular actors, I interpret these as twenty-first-century extensions of Sterne's 'audile technique'.

The implementation and widespread adoption of Facebook's autoplay standard set a particular trans-platform norm – of default on-vision and default off-sound. This norm – and, importantly, an understanding of it *as* a norm, and a kind of vernacular theorizing in response to that norm – led to the proliferation of new media behaviours, concretized in digital artefacts and genres. One such set of artefacts is what I have elsewhere called the 'unmute this' genre of video posts.[7] These were posts – often on the platforms Tumblr or Twitter – with a video whose content might seem to be fully comprehensible with visuals alone (i.e., muted), but to which had been appended a caption entreating a passing viewer to 'unmute', to turn the sound on for a fuller experience (Figure 2). For example, a video of a dancing dog which, when unmuted, turns out to be 'dancing' to the instrumental opening to Toto's 'Africa', a llama whose leaps across a field are, with the sound on, choreographed to the chorus of DMX's 'Party Up', or an adorable baby sloth who makes an even more adorable noise when unmuted.[8]

On such videos, the directive of 'unmute this' captions functions as a comprehensible and personal communication from and to savvy users. This genre of post has largely been produced and circulated on platforms that enable nested content – a post within a post – where instructional captions such as 'sound on!' or 'unmute, i am begging you' can come from someone other than the initial creator or poster of the video. 'Unmute this' video posts rely on trust, a system of user-to-user curation of (and vouching for) a worthwhile trade-off between effort and pleasure.[9] This is one

6 Jonathan Sterne, *The Audible Past: Cultural Origins of Sound Reproduction* (Durham, NC: Duke University Press, 2003).

7 Harper, '"Unmute This"'.

8 Many, though not all, 'unmute this' videos feature animals, either with their own surprising sounds or with humorous added soundtracking. For analysis and implications of online animal musicking, see Kate Galloway's contribution to this issue, 'Sensing, Sharing, and Listening to Musicking Animals across the Sonic Environments of Social Media', in this issue.

9 This understanding parallels what Raquel Campos Valverde calls 'imagined listening' and 'imagined audiences' of music shared on social media; 'unmute' directives speak more to imaginings of listeners' media ecosystems and

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

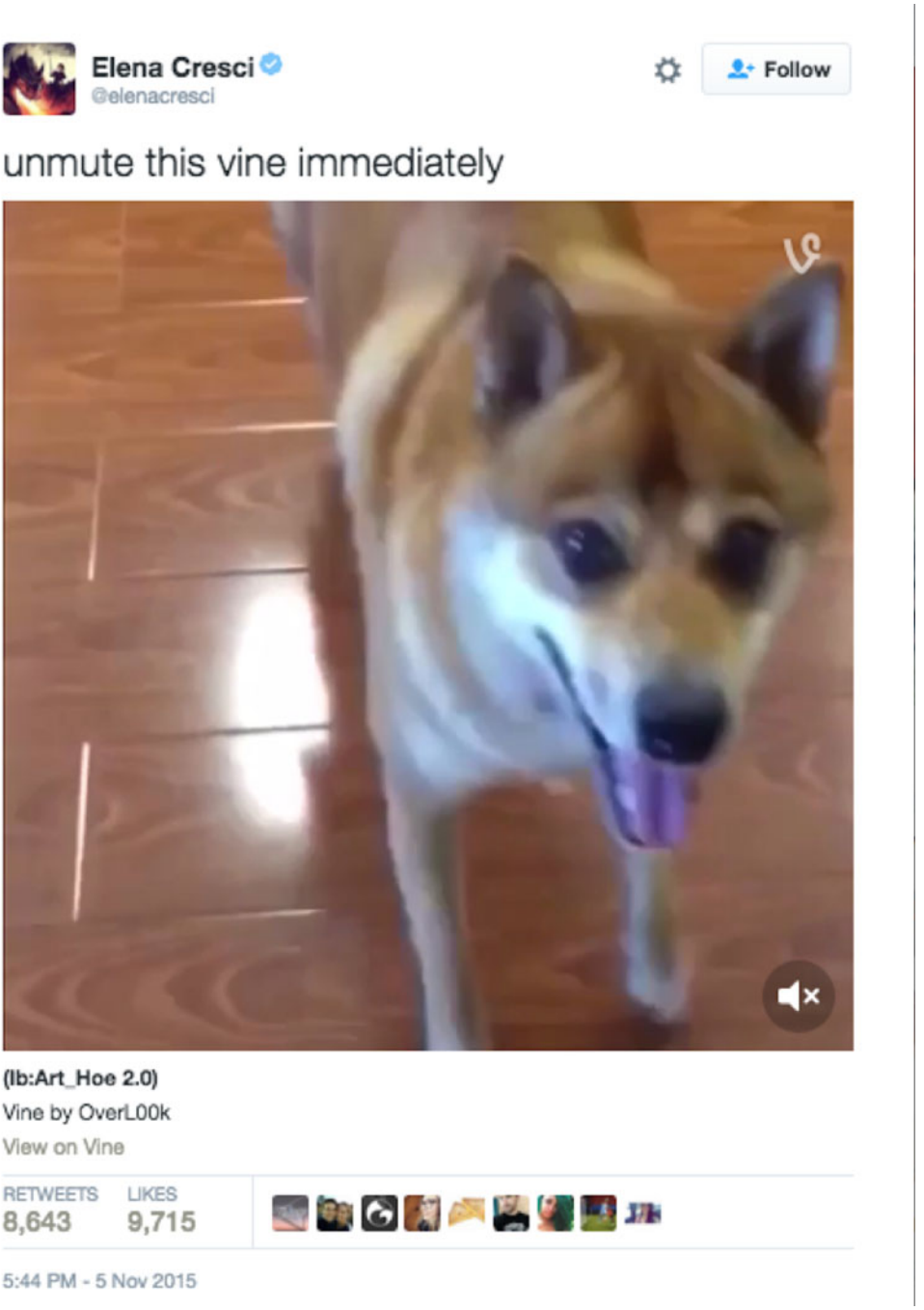


Figure 2 (Colour online) Screenshots of social media posts with embedded videos, all featuring captions directing viewers to unmute (2014–16).

form, rather than to musical content and listeners' experiences thereof. Raquel Campos Valverde, 'Online Musicking for Humanity: The Role of Imagined Listening and the Moral Economies of Music Sharing on Social Media', *Popular Music* 41/2(2022), https://doi.org/10.1017/S0261143022000034.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

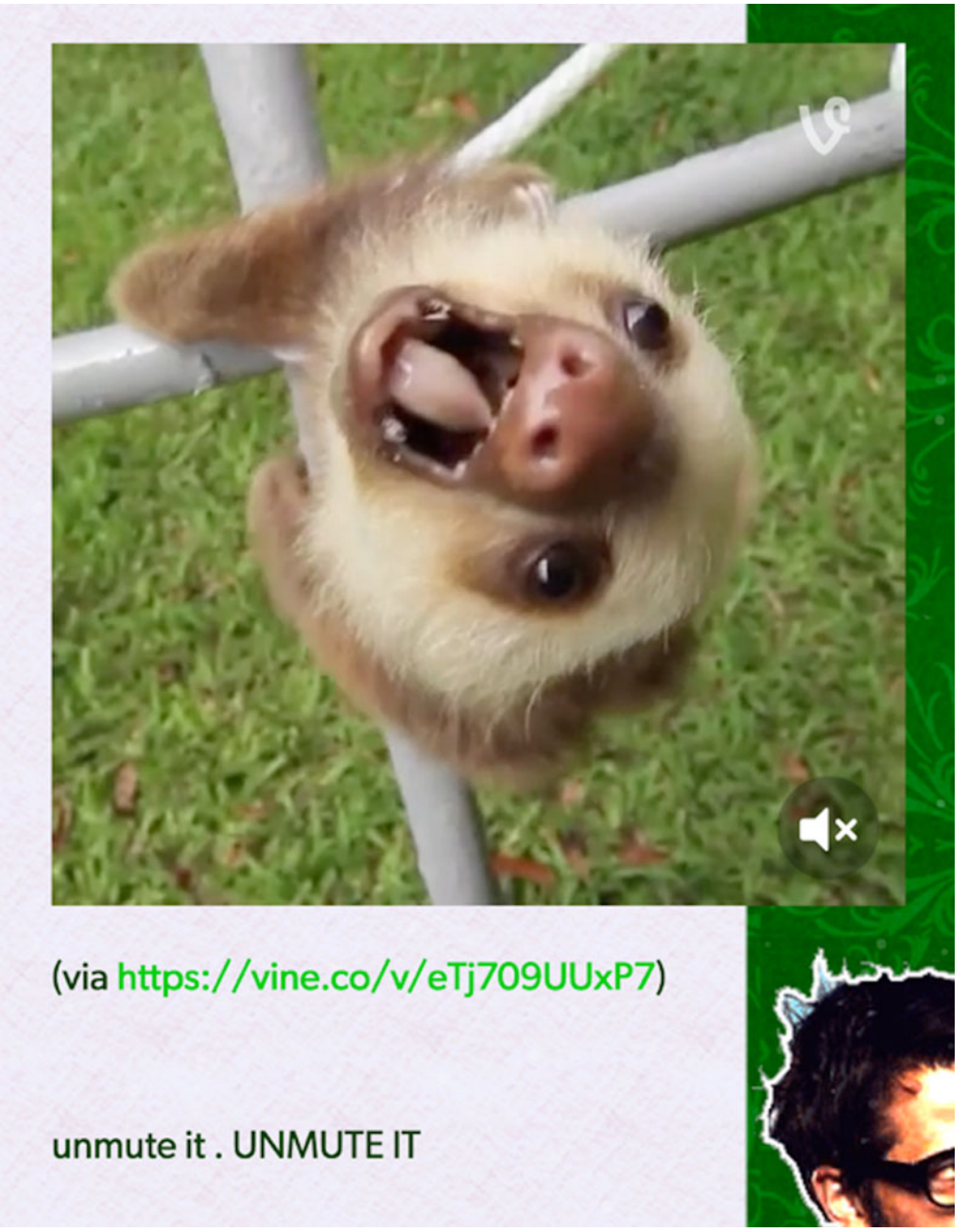
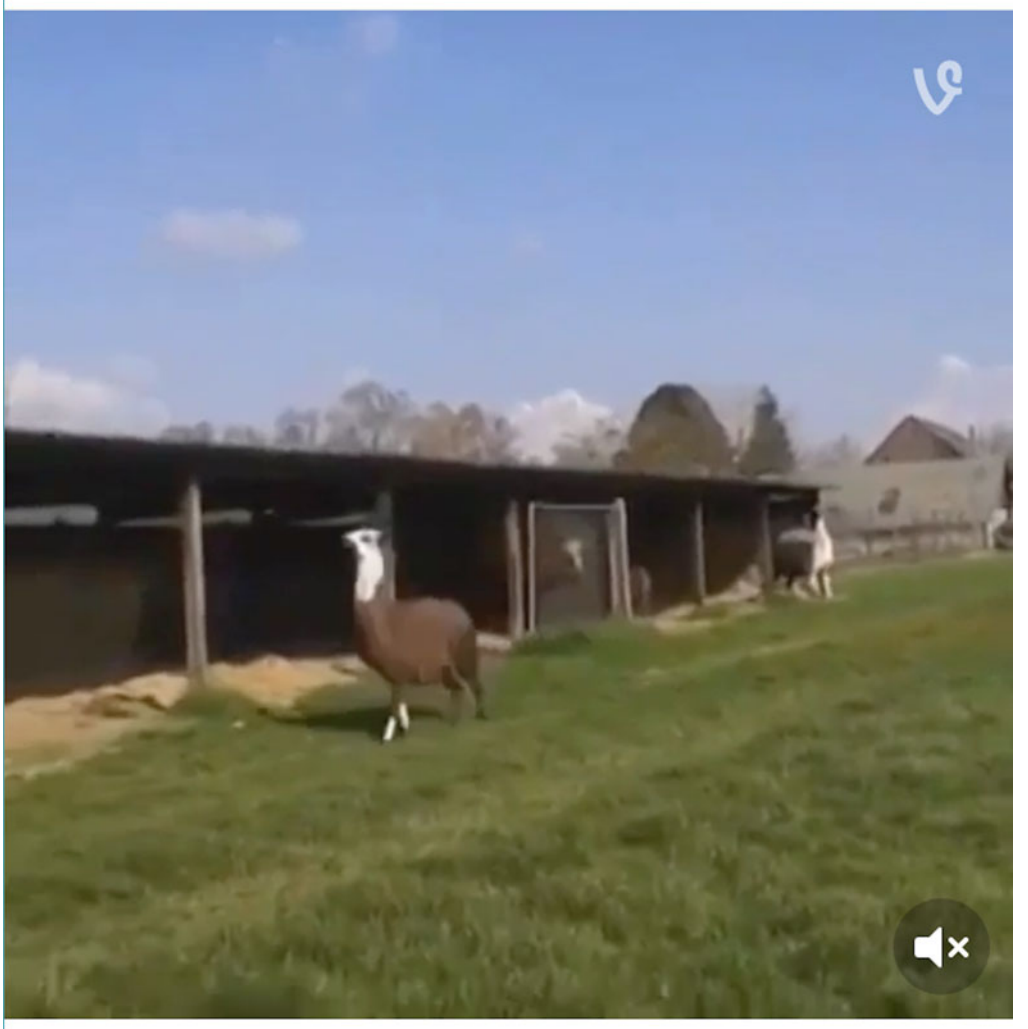


Figure 2 Continued.

major implication of the 'unmute this' video post genre that interests me – users contributing to posts to curate and afford an ideal content experience, based on assumptions of particular media conditions.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press



I wanna live like this llama

Please unmute this.

Source: vine.co

314,859 notes Oct 7th, 2014

Figure 2 Continued.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

What are these assumed conditions? Users are assuming the multi-platform muted-autoplay standard established by Facebook in the mid-2010s and subsequently normalized beyond it. The assumed condition might also be Michael Bull's sonically 'cocooned' headphone user – in which a user ensconces themself in a curated sonic world technologically separated from the aural stimuli occurring around them.[10] More broadly, one might read the assumption behind 'unmute this' directions as an imagining of the kind of ubiquitous listener that Anahid Kassabian theorizes, in which 'we listen "alongside," or simultaneous with, other activities'. In this understanding, the 'effort' to comply with an 'unmute' directive might be understood to involve a multi-step process of navigating between a variety of mobile hardware and software, switching apps, putting on headphones, and so on.[11]

Further, the audile techniques made manifest in 'unmute this' directives assume users who are most paradigmatically urban-dwelling and upper-middle class, expanding only slightly from the white-collar bullpen- or cubicle-dwellers of the anti-autoplay tech journalism pieces cited previously. Contestations around what kinds of people (and what kinds of music and sounds) are allowed to take up public or shared space often fracture along lines of racial and socioeconomic difference, as well, and understanding a Michael Bull-esque 'sonic cocoon' to be the 'appropriate' techno-sonic mode of being and moving in the world is, of course, not a universal, but instead a particular hegemonic default.[12] Byrd McDaniel, in his contribution to this issue, analyses music podcasts as 'persuasive demonstrations of music consumption', and demonstrates how they model particular 'correct' and 'expert' modes of listening that encode a normative white, middle-class point of audition. As these podcasts model particular listening modes via their content, these protocols and post genres model assumed listening techniques – and assumed listeners – via their format.[13]

There is an interesting flip, in the 'unmute' directives, of what Sterne, in the introduction to *The Audible Past* calls the 'audiovisual litany' – binarized assumptions that tend to frame contemporary Western understandings of the sense of hearing, arranged in opposition to vision. Included in this litany are the tenets that hearing is immersive and spherical, while vision is superficial, perspectival, and directional, and that hearing is an 'always on' sense, while vision is elective.

But 'unmute this' directives rely on precisely the opposite rationale. Rather than an assumption of continuous, automatic *listening* and elective *vision*, in the 'unmute this' directive, users assume an ecosystem in which vision is always being passively engaged, but hearing is marked and discretionary, and must be actively singled out. One's fullest attention, it is

10 Michael Bull, *Sound Moves! iPod Culture and Urban Experience* (New York: Routledge, 2007).

11 Anahid Kassabian, *Ubiquitous Listening: Affect, Attention, and Distributed Subjectivity* (Berkeley, CA: University of California Press, 2013), 9.

12 See Isabelia Herrera, 'When a Minivan becomes a Music Machine', *New York Times*, 21 September 2021, www.nytimes.com/interactive/2021/09/21/arts/music/dominican-soundsystems.html; Brandi Thompson Summers, 'Reclaiming the Chocolate City: Soundscapes of Gentrification and Resistance in Washington, DC.', *Environment and Planning D: Society and Space* 39/1 (2021); Brian Larkin, 'Techniques of Inattention: The Mediality of Loudspeakers in Nigeria', *Anthropological Quarterly* 87/4 (2014).

13 Byrd McDaniel, 'All Songs Considered: The Persuasive Listening of Music Podcasts', this issue.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

assumed by these users, is a listening, rather than viewing, attention – or rather, it is a fully audiovisual attention.

The word 'attention' invokes what has been dubbed 'the attention economy' by theorists such as Michael Goldhaber, Jonathan Beller, Richard Lanham, and Tim Hwang.[14] This notion of the attention economy, one which has been increasingly deployed as a diagnostic and heuristic for analysing digital behaviour, apparatuses, and phenomenology, posits a system in which information is overabundant, and attention is in a relative state of scarcity – and therefore it can be ascribed value and can function as a commodity. While theorists such as Lanham make the point that provoking, curating, and capturing attention have long histories in theories of aesthetics and rhetoric, the twenty-first-century digital ecosystem of apps and platforms, networked digital devices, and digital identities has spectacularized information overabundance and the understood attendant crises of 'attention', often framed in terms of sensory perception and its capacities and limits.

'Unmute this' directives, as I have argued, operate within a system of social credentialing, with trust in any particular user's 'unmute!' exclamations balancing out a perceived risk or inconvenience – taking the time to direct sole focus to a particular platform and object, with digital attention figured as a zero sum game. The risk is therefore related to the commodity status of attention, and the user's acquiescence in attention and time are understood to have value, to be transmutable to other more literal forms of capital. In the heyday of 'unmute this' videos and posts across the mid-2010s, it was also easy to find accompanying expressions of exhaustion by users frustrated with the perceived overabundance of such tags (Figure 3).

The frustration communicated in those screenshots conveys an implicit breach of trust in lack of payoff for individual videos. It could also be seen as concern for the ongoing indicative functionality of 'unmute this' captions – that is, urging the perceived non-normative user behaviours of turning sound on and watching video in its temporal and audiovisual entirety – if too broadly applied. Also latent is the recognition of the tagging and captioning process as an exploitable feature, and user awareness of the value of their behaviours and perception as harvestable attention-data.

Captions, cooption, commodification

If 'unmute this' directives and tags constituted a vernacular response to the implementation and standardization of (muted) autoplaying video functionality, these expressions of frustration about an 'overuse' of 'unmute' directives point to the (inevitable) cooption of the format.

14 See Michael Goldhaber, 'The Attention Economy and the Net', *First Monday* 2/4 (1997), http://firstmonday.org/ojs/index.php/fm/article/view/519/440; Jonathan Beller, *The Cinematic Mode of Production: Attention Economy and the Society of the Spectacle* (Hanover, NH: Dartmouth College Press, 2006); Richard Lanham, *The Economics of Attention: Style and Substance in the Age of Information* (Chicago: University of Chicago Press, 2006); Tim Hwang, *The Subprime Attention Crisis: Advertising and the Time Bomb at the Heart of the Internet* (New York: FSG Originals X Logic, 2020). Yves Citton's *The Ecology of Attention* presents a valuable critique of the attention economy framing, urging theorists of attention and sensory perception to expand metaphorical models beyond the economic to the interactional and ecological. Citton's emphasis on the relational and collective nature of attention is an important aspect for scholars working within digital platform capitalism to consider. See Yves Citton, *The Ecology of Attention* (New York: Polity, 2017).

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

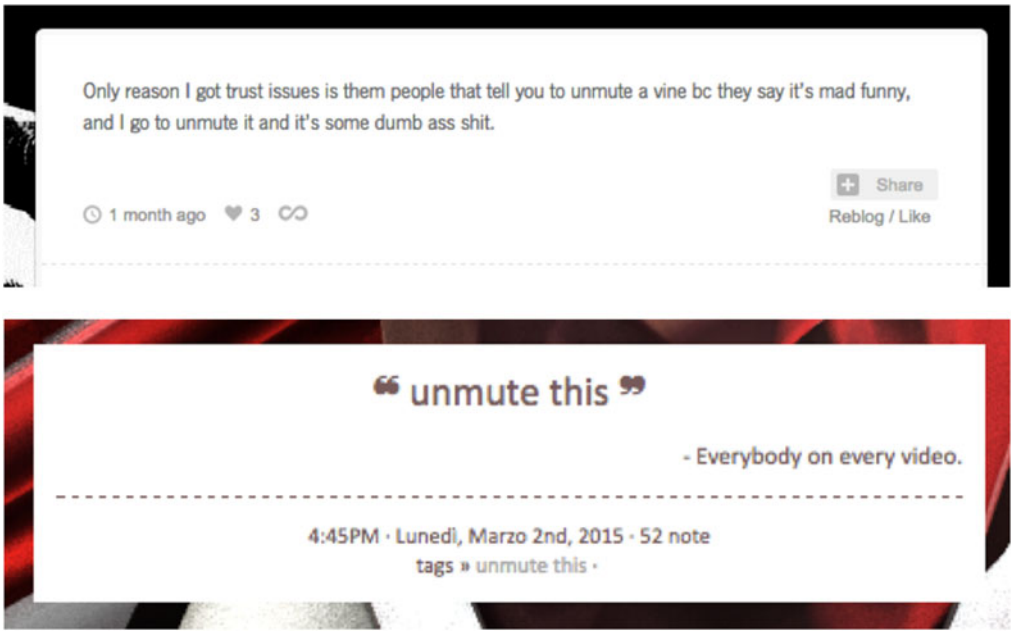


Figure 3 (Colour online) Social media posts critiquing the omnipresence of 'unmute this' captions (2015–16).

While the preceding frustrated posts mostly directed their ire at other individual content creators, users finding ways to playfully engage with autoplay was immediately advantageous for large platforms; through such behaviour, users normalize and habituate to the format. A few examples of the reification of this technique follow: corporate strategies to monetize unmuting illustrate an understanding of the behaviour as valuable in and of itself.

At a somewhat foundational functional level, an update to Google Chrome's API in April 2018 selectively allowed web autoplay of particular audiovisual objects on websites based on a user's 'Media Engagement Index' – a quantified, archived history of behaviour – with that page.[15] Drawing even more closely on the legibility of 'unmute this' videos and similar phenomena, platforms such as Instagram and Snapchat developed top-down functionalities to encourage the production of 'sound on' directives by their users – the prominent inclusion of speaker icon emojis or stickers that users could easily append to their posts, and that, in some instances, would themselves function as toggles to switch user sound on (Figure 4).

For Instagram, this implementation was directly beneficial to advertisers – turning on sound for one video in a feed turns on sound for all subsequent videos, including targeted advertiser posts automatically inserted into a user's stories and feed. These examples further demonstrate how 'attention' in the platform assemblage of current social and digital media is understood as being sonic, based in sound and listening, or as being fully audiovisual. As

15 See François Beaufort, 'Autoplay Policy in Chrome', *Chrome Developers Blog*, 13 September 2017, https://developer.chrome.com/blog/autoplay/.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

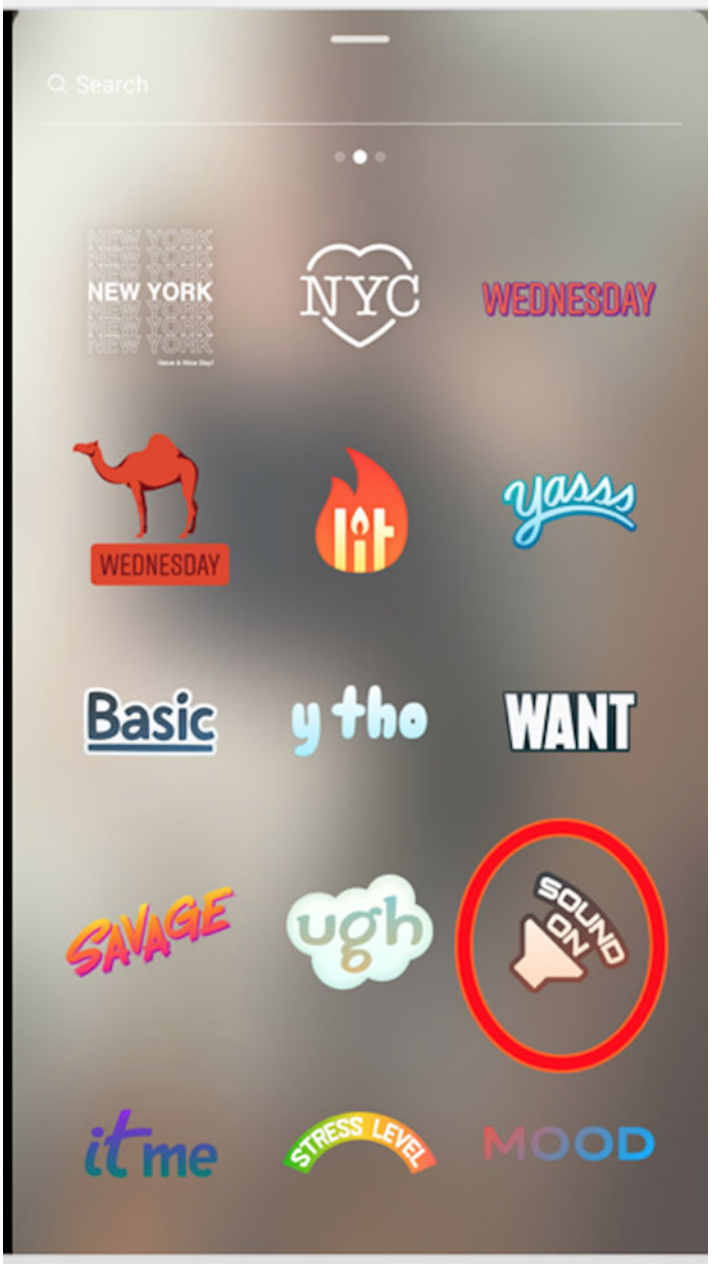


Figure 4 (Colour online) 'Sound On' icons as easily addable stickers on Instagram and Snapchat (2018).

media theorist Taina Bucher argues, 'attention governs and is governed in concrete material contexts and assemblages. Protocols, algorithms and buttons do not merely mediate modes of paying attention but also shape the conditions of the sensible.'[16]

16 Taina Bucher, 'A Technicity of Attention: How Software "Makes Sense"', *Culture Machine* 13 (2012), 17.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

Understandings of 'Sound-on'-ing as audile technique become the basis of corporate attention economy strategy, rhetoric, and iconicity. That said, sonic *in*attention, too, has been leveraged as a commodity in the post-2010s ecosystem of digital platforms and devices. Anthropologist and media theorist Brian Larkin has productively demonstrated that inattention is not simply attention's failed opposite, as preceding conceptions framed it – '[i]nattention is not not attending'.[17] Instead, inattention can itself be a cultivated skill, a set of bodily and perceptual techniques honed in response to a particular medial environment. The mediated, afforded option to easily exercise that set of techniques – digital platforms concretize them into platform features and charge a premium for making them easily accessible. One such example is the YouTube Red service. YouTube Red launched in 2016 as a rebranded and rethought version of ongoing efforts to monetize the platform as a subscription service, as well as to shift the platform's identity, content, and revenue model to more closely parallel that of other proprietary media services such as those of Amazon, Netflix, and Disney. While a central selling point of the subscription service was the removal of advertisements autoplaying before, during, and after videos, YouTube Red (since further rebranded in 2018 as 'YouTube Premium') also allowed for listening to videos with the screen off, or with the app minimized (applicable to mobile devices only).[18] While standard YouTube viewing attempts to direct and monetize attention through ad placement, this feature monetizes a different set of perceptual preferences – which are understood to be not just passive, but actively desired and engaged. YouTube Red users quite literally pay for inattention as a premium platform feature – with 'attention' figured as an onerously engaged audiovisuality – along with its attendant implications of a variety of pragmatics such as increased battery usage, singular platform or app focus, or a necessarily more careful handling of the device.

While responses to Facebook's autoplay standards evolved in a variety of distinct ways across the other major platforms mentioned in this article, Facebook itself elevated a different norm within its own ecosystem for dealing with default sound-off. While some 'Sound-on'-ing exists in the Facebook platform environment, the style of videos there, in the post-autoplay feed landscape, have more consistently foregrounded the expected user interaction conditions of muted sound and short engagement periods, with the use of lots of onscreen text and 'front-loaded' content meant to transmit as much material as possible in the first seconds of viewing. Contributing to this standard, Facebook highly promoted the implementation of closed-captioning features; while third-party software for adding captions to digital video on other platforms existed as an option prior to the social media autoplay renaissance, Facebook began offering its own in-house captioning as an option for the platform's US users from 2014, and then providing it automatically for its advertisers from early 2016 onwards.

17 Larkin, 'Techniques of Inattention', 1008. While the stakes of inattention are far more heightened and visceral for Larkin's Nigerian interlocutors than for most Spotify or YouTube subscribers, I find his intervention valuable for considering inattention as something other than attention's inert opposite.

18 See Ben Popper, 'Red Dawn', The Verge, 21 October 2015, www.theverge.com/2015/10/21/9566973/youtube-red-ad-free-offline-paid-subscription-service.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

The introduction of captions was framed as an accessibility feature by Facebook, and celebrated as such by accessibility activists and those benefiting from the newly proliferating modality. However, the feature introduction was also discursively framed in terms of attention economy scarcity. In one representative tech blog post, the suggestion is that the appropriation of an accessibility feature in its 'present' usage eclipses closed captioning's accessible 'past':

> Commonly used in the past to aid those with hearing disabilities, closed captions have been given new life as a tool for the distracted and time-strapped . . . If you upload subtitles, Facebook will automatically serve them with your video whenever it is muted. Once unmuted, they disappear.[19]

Following Facebook's promotion of captioning, other video-hosting platforms and applications, from Instagram to TikTok, have only amplified and accelerated the use of functionalities such as captioning and text-to-speech. Disability scholars such as Mara Mills and Jay Dolmage encourage analysis of the use of these functionalities as a 'genre of ekphrasis' – originally, the artistic translation of visual material into verbal.[20] In this mode, we might comprehend captions, transcriptions, image descriptions, or speech-to-text overlay as 'adding artistic and rhetorical value, not simply transposing or distilling meanings'.[21] Certainly, on platforms such as TikTok, where sound-on is the norm, video creators nonetheless frequently deploy on-screen captioning or make use of speech-to-text functionalities: both within the platform and parasitic to it, whole hosts of options are available that fuse the aesthetic with the accessible, resulting in video texts with several reinforcing-but-not-exactly-redundant layers. Users on these platforms creatively incorporate and manipulate multiple planes of meaning-making, resulting in sharable video texts optimized in a variety of ways – for accessibility, for shareability across various platforms (and platform sound-on norms), for consumption in a variety of modalities, and for humour and aesthetics.

On the one hand, the proliferation of captioning and related features as ubiquitous digital norms has made even the most odd and ephemeral internet content more broadly accessible to users with a range of sensory impairments and abilities. The mass adoption of these features is also potentially deeply lucrative – and therefore highly encouraged and sought after – by corporate interests seeking to collect and parse sound as a relatively unexploited data frontier. As Jonathan Sterne and Elena Razlogova note in an article on the use of artificial intelligence in audio mastering:

> The neural networks through which machine learning applications run require vast amounts of data. These data have to be generated by people or organizations, often for free or as a side-effect of other processes, and they have to be in a form available

19 Luke Dringoli, 'How to Boost Facebook Engagement Using Video Subtitles', *MediaCause.com*, 16 June 2016, https://mediacause.com/create-add-facebook-video-subtitles/.

20 Mara Mills, 'Listening to Images: Audio Description, the Translation Overlay, and Image Retrieval', *The Cine-Files* 8 (Spring 2015); Jay Dolmage, *Disability Rhetoric* (Syracuse, NY: Syracuse University Press, 2014).

21 Mills, 'Listening to Images', 2.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

> for processing . . . In other words, they are only available because of a more fundamental infrastructural system where a wide range of commensurable data can be collected from people and then transmitted, circulated, collected, and repurposed.[22]

Sterne and Razlogova acknowledge that '[i]n the case of audio, the vast number of digital recordings in existence would seem to make it a perfect resource for training and testing machine learning applications. But sound – especially the sound of music – has been a difficult area for artificial intelligence.'[23] Voluntary participation in the use of captioning, speech-to-text, and other similar softwares – and, in some instances, the uncompensated labour of providing, proofreading, or improving the performance of those softwares – contributes to the generation of the necessary datasets, in forms useful to the corporate interests that hold and manage them, and parse them for re-sale use in advertising and other guises.[24]

Conclusion: Sound On

The Pandora 'Sound On' campaign – the 2018 and 2019 multi-city 'brand platform' launch that caught my attention in the opening anecdote – draws from precisely the same catchy construction as the preceding Instagram and Snapchat 'sound on' examples, adopting the diction of vernacular sociality in an attempt to encourage users to 'live their lives at full volume'. It is no coincidence that the Pandora campaign was directly targeted at sites of very literal mobility and urban circulation – with many of the ads directly placed in targeted areas of travel and commute: subways, airports, and train and bus terminals.[25] Through word and iconography choice, the advertising of the campaign evoked vernacular 'sound on' and 'unmute' directives – even though the targeted Pandora users in question would most likely be using the platform's sounds in a more standard 'background listening' mode, to cocoon themselves from the sonic chaos and unpredictability of the urban commute, within a musical world of their own making. However, by drawing in a user captioning practice for denoting especially attention-worthy autoplaying video, the Pandora ad campaign elides the more standard use case of its platform with indicators for full digital audiovisual attentiveness, bringing its background music commodity to the foreground.

In this article, I hope to have shown how negotiations of sound and listening have factored deeply into a variety of contemporary attempts to harness and monetize attention as a commodity in a digital economy of platforms, advertisements, and data. The digital attention

22 Jonathan Sterne and Elena Razlogova, 'Tuning Sound for Infrastructures: Artificial Intelligence, Automation, and the Cultural Politics of Audio Mastering', *Cultural Studies* 35/4–5 (2021), 754. See also danah boyd and Kate Crawford, 'Critical Questions for Big Data', *Information, Communication & Society* 15/5 (2012); Kate Crawford and Vladan Joler, 'Anatomy of an AI System: The Amazon Echo as an Anatomical Map of Labor, Data and Planetary Resources', *AI Now Instititute and Share Lab*, 7 September 2018, https://anatomyof.ai.

23 Sterne and Razlogova, 'Tuning Sound for Infrastructures', 754.

24 This is a process paralleling what Stephen Neville calls 'eavesmining' by virtual household 'assistant' appliances such as the Amazon Alexa. See Stephen Neville, 'Eavesmining: A Critical Audit of the Amazon Echo and Alexa Conditions of Use', *Surveillance and Society* 18/3 (2020).

25 Katie Richards, 'Pandora Plans to Invade Your Holiday Travel with the Launch of "Sound On"', *Adweek.com*, 8 November 2018, www.adweek.com/brand-marketing/pandora-plans-to-invade-your-holiday-travel-with-the-launch-of-sound-on/.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

economy, I stress, is fundamentally invested in the aural as well as the ocular, far beyond specifically musical spaces. Reading across a variety of platforms, I have suggested ways in which the powerful ubiquity of a particular functionality – in this case, autoplaying video – can underpin a wide variety of responses – techniques, formats, behaviours, genres – all built around particular shared assumptions regarding sensory techniques. Those various responses – in particular the corporate cooption of vernacular practices such as 'unmute this' tags – suggest the understood power of sound (and behaviours around sound, hearing, and listening) as a parameter of 'attention' in the digital ecosystems of the twenty-first century: as a structuring component of platform protocols, functionalities, and behaviours, and as a site of contestation, commodification, and creativity.

Bibliography

Beaufort, François. 'Autoplay Policy in Chrome'. *Chrome Developers Blog*. 13 September 2017. https://developer.chrome.com/blog/autoplay/ (accessed 22 July 2022).

Beller, Jonathan. *The Cinematic Mode of Production: Attention Economy and the Society of the Spectacle*. Hanover, NH: Dartmouth College Press, 2006.

Blattberg, Eric. 'Like It or Not, Autoplay Video Won'. *Digiday*. 21 April 2015. http://digiday.com/publishers/autoplay-video-beat-regular-video-sorry-guys/ (accessed 22 July 2022).

boyd, danah and Kate Crawford. 'Critical Questions for Big Data'. *Information, Communication & Society* 15/5 (2012), 662–79.

Bucher, Taina. 'A Technicity of Attention: How Software "Makes Sense"'. *Culture Machine* 13 (2012), 1–23.

Bull, Michael. *Sound Moves! iPod Culture and Urban Experience*. New York: Routledge, 2007.

Campos Valverde, Raquel. 'Online Musicking for Humanity: The Role of Imagined Listening and the Moral Economies of Music Sharing on Social Media'. *Popular Music* 41/2 (2022), 194–215. https://doi.org/10.1017/S0261143022000034.

Citton, Yves. *The Ecology of Attention*. New York: Polity, 2017.

Crawford, Kate and Vladan Joler. 'Anatomy of an AI System: The Amazon Echo as an Anatomical Map of Labor, Data and Planetary Resources'. *AI Now Insitiute and Share Lab*. 7 September 2018. https://anatomyof.ai (accessed 22 July 2022).

Dolmage, Jay. *Disability Rhetoric*. Syracuse, NY: Syracuse University Press, 2014.

Dringoli, Luke. 'How to Boost Facebook Engagement Using Video Subtitles'. *MediaCause.com*. 16 June 2016. https://mediacause.com/create-add-facebook-video-subtitles/ (accessed 22 July 2022).

Galloway, Kate. 'Sensing, Sharing, and Listening to Musicking Animals across the Sonic Environments of Social Media.' This issue.

Goldhaber, Michael. 'The Attention Economy and the Net'. *First Monday* 2/4 (1997). http://firstmonday.org/ojs/index.php/fm/article/view/519/440 (accessed 22 July 2022).

Harper, Paula. '"Unmute This": Captioning an (Audio)Visual Microgenre'. *The Soundtrack* 9/1–2 (2016), 7–23.

Hazard Owen, Laura. 'Did Facebook's Faulty Data Push News Publishers to Make Terrible Decisions on Video?'. *NiemanLab.org*. 17 October 2018. www.niemanlab.org/2018/10/did-facebooks-faulty-data-push-news-publishers-to-make-terrible-decisions-on-video/ (accessed 22 July 2022).

Herrera, Isabelia. 'When a Minivan becomes a Music Machine'. *New York Times*. 21 September 2021. www.nytimes.com/interactive/2021/09/21/arts/music/dominican-soundsystems.html (accessed 22 July 2022).

Hwang, Tim. *The Subprime Attention Crisis: Advertising and the Time Bomb at the Heart of the Internet*. New York: FSG Originals X Logic, 2020.

Kassabian, Anahid. *Ubiquitous Listening: Affect, Attention, and Distributed Subjectivity*. Berkeley, CA: University of California Press, 2013.

Lanham, Richard. *The Economics of Attention: Style and Substance in the Age of Information*. Chicago: University of Chicago Press, 2006.

Larkin, Brian. 'Techniques of Inattention: The Mediality of Loudspeakers in Nigeria'. *Anthropological Quarterly* 87/4 (2014): 989–1015.

Mayes, Kelly. 'An Easier Way to Watch Video'. *Facebook Newsroom*. 12 September 2013. https://about.fb.com/news/2013/09/an-easier-way-to-watch-video/ (accessed 22 July 2022).

McDaniel, Byrd. 'All Songs Considered: The Persuasive Listening of Music Podcasts'. This issue.

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press

Mills, Mara. 'Listening to Images: Audio Description, the Translation Overlay, and Image Retrieval'. *The Cine-Files* 8 (Spring 2015), 1–5.
Neville, Stephen J. 'Eavesmining: A Critical Audit of the Amazon Echo and Alexa Conditions of Use'. *Surveillance and Society* 18/3 (2020), 343–56.
Popper, Ben. 'Red Dawn'. *The Verge*. 21 October 2015. www.theverge.com/2015/10/21/9566973/youtube-red-ad-free-offline-paid-subscription-service (accessed 22 July 2022).
Richards, Katie. 'Pandora Plans to Invade Your Holiday Travel with the Launch of "Sound On"'. *Adweek.com*. 8 November 2018. www.adweek.com/brand-marketing/pandora-plans-to-invade-your-holiday-travel-with-the-launch-of-sound-on/ (accessed 22 July 2022).
Sternberg, Josh. 'The Most Hated Digital Ad Tactic'. *Digiday*. 3 April 2013. https://digiday.com/media/the-most-hated-digital-ad-tactic/ (accessed 22 July 2022).
Sterne, Jonathan. *The Audible Past: Cultural Origins of Sound Reproduction*. Durham, NC: Duke University Press, 2003.
Sterne, Jonathan and Elena Razlogova. 'Tuning Sound for Infrastructures: Artificial Intelligence, Automation, and the Cultural Politics of Audio Mastering'. *Cultural Studies* 35/4–5 (2021), 750–70.
Summers, Brandi Thompson. 'Reclaiming the Chocolate City: Soundscapes of Gentrification and Resistance in Washington, DC'. *Environment and Planning D: Society and Space* 39/1 (2021), 30–46.
Vranica, Suzanne, and Jack Marshall. 'Facebook Overestimated Key Video Metric for Two Years'. *Wall Street Journal*. 22 September 2016. www.wsj.com/articles/facebook-overestimated-key-video-metric-for-two-years-1474586951 (accessed 22 July 2022).

https://doi.org/10.1017/S1478572222000287 Published online by Cambridge University Press